



Early Journal Content on JSTOR, Free to Anyone in the World

This article is one of nearly 500,000 scholarly works digitized and made freely available to everyone in the world by JSTOR.

Known as the Early Journal Content, this set of works include research articles, news, letters, and other writings published in more than 200 of the oldest leading academic journals. The works date from the mid-seventeenth to the early twentieth centuries.

We encourage people to read and share the Early Journal Content openly and to tell others that this resource exists. People may post this content online or redistribute in any way for non-commercial purposes.

Read more about Early Journal Content at <http://about.jstor.org/participate-jstor/individuals/early-journal-content>.

JSTOR is a digital library of academic journals, books, and primary source objects. JSTOR helps people discover, use, and build upon a wide range of content through a powerful research and teaching platform, and preserves this content for future generations. JSTOR is part of ITHAKA, a not-for-profit organization that also includes Ithaka S+R and Portico. For more information about JSTOR, please contact support@jstor.org.

A CRITIQUE OF PSYCHO-PHYSIC METHODS.

BY JOSEPH JASTROW, PH. D.

Probably no result of experimental psychology has been cited as frequently and with as much confidence in its truth and importance as the psycho-physic law. By some it has been regarded as worthy of ranking with the law of gravitation, while others refuse to recognize it at all. A clear statement of what the law means is rarely found ; sometimes the term refers to the results of Weber's experiments, and again it refers to Fechner's deductions concerning the relation of stimulus to sensation. But these two are in a sense totally distinct, and should be kept so. The psycho-physic methods are applicable only to such experiments as can be utilized for establishing Weber's law. And this paper is to be devoted to a rigorous logical criticism of the methods and interpretation of such psycho-physic experiments. Its object is a practical one: to clear the way for a more rational system of psycho-physics by directing future experimentation into that path in which it is most promising of results, and thus preventing the employment of the many uncritical and unanalyzed processes now current.

Where the falsity of one point is so closely connected with the falsity of many another, it is difficult to know where to begin and how to proceed. The full appreciation of one point requires a knowledge of the con-

siderations that follow ; hence the order of exposition will have to follow convenience rather than logical sequence.

THE THREE PSYCHO-PHYSIC METHODS.

And first let us state briefly what the three usually recognized psycho-physic methods are :¹

I. *The Method of the Just Observable Difference.*—This is the method that Weber followed in his celebrated experiments, and has, I believe, done much to introduce radical misconceptions into psycho-physics. This method consists in applying a certain stimulus to the sensitive surface of the subject, and then finding the least greater or the least smaller stimulus which can just be recognized as different ; one either adds small increments to the initial stimulus until the stimulus thus formed is felt to be greater, or lessens the intensity of the initial stimulus until the diminution is clearly noticed, and notes in either case the ratio of the alteration to the original stimulus. It is considered best to employ both these processes, and regard the mean of the two results thus obtained as the true just observable difference. The ratio of the difference between the initial and the altered stimulus to the initial stimulus, or better, to half the sum of the two, measures the differential sensibility. The smaller this ratio the finer the sensibility is said to be.

II. *The Method of Right and Wrong Cases.*—One here chooses two slightly different stimuli and presents them to the sensitive surface of the subject, requiring him to judge whether the first stimulus is

¹ Other proposed psycho-physic methods, and especially that of the "mean gradations" (*mittlere Abstufungen*), deserve a separate treatment ; this I hope to furnish on another occasion.

greater or less than the second, and records the number of cases in which the decision is right and the number in which it is wrong. The ratio of right answers to the total number of answers measures the sensibility and varies in a direct sense with it.¹

III. *The Method of the Average Error* (or as I prefer to call it, of the Probable Error).—In this method a stimulus is presented, and the subject is required to adjust a second stimulus so as to be equal to the first. The average deviation of the several adjustments from their mean (or better, the probable error of the adjustments) directly measures the sensibility.

CRITIQUE OF METHOD I.

Applying the method of the just observable difference to pressure sensations, Weber was led to formulate his well known law. In this he announced that if you apply a certain weight to the skin and find the least greater (or least smaller) weight which can be recognized as different, and then take an entirely different weight and repeat the process, the ratio of the first to the second of each pair of weights used in any such experiment will be the same ; that is, the ratio of the just observable difference to the initial weight is a constant. This process simply compares the recognition of a difference in one part of the psychic scale with that in another ; it says nothing and cannot be made to say anything about the ratio of stimulus and sensation. It is not a psycho-physic, but a psycho-psychic, law. So much is plain.

But the experiment of Weber is in every way vague and inexact. To begin with, the bare statement that

¹This is the usual form of the method. Variations in this proceeding will appear later.

one pressure is "just distinguishable" from another is altogether indefinite. This expression may have as many as four practically distinct meanings. It may refer (1) to two pressures sufficiently different to enable the subject *sometimes* to tell which is which; or (2) to differences that will *always* be correctly recognized; or (3) such as will only accidentally fail to be recognized; or finally (4) it may refer to differences which will be correctly judged any characteristic proportion of times between these extremes. Now the just observable difference will evidently be a totally different thing according to which of these interpretations is chosen. The first interpretation has rarely if ever been used, because in practice any difference that is likely to occur, however minute, will be sometimes correctly appreciated, and will sometimes arouse a confidence in the recognition of it sufficient to hazard an answer upon; yet it has quite as much in its favor as any other interpretation, and leads to equally definite conceptions. The second is the method usually chosen, but is almost sure to degenerate into the third. Under the fourth head any difference of stimuli correctly appreciated any given ratio of times may be used; the simplest is that in which in half the trials the difference is correctly judged, and this interpretation has found favor with some workers.

A very brief consideration of the psychological processes involved in such judgments is sufficient to show that in practice none of the four interpretations is useful or valid. The answer which the subject gives may be of two kinds: (*a*) if he is asked to mention *when* he feels disposed to regard the sensations caused by the two stimuli as different (and probably, too, noting the direction of this difference), his answer depends mainly, if

not entirely, on the minimum amount of confidence upon which he finds himself disposed to make a judgment; (b) if he is asked to decide which he judges to be the more and which the less intense stimulus—in other words, if the correctness of the answer (made on the basis of a certain stimulus) is the deciding point and not the disposition to answer, then we are simply using the method of the right and wrong cases in a rather loose form. As a matter of fact, the just observable difference method pure and simple has seldom if ever been used; the correctness of the answer is always to some extent taken into account and in this form the method loses all *raison d'être*.

Again, if we take Fechner's definition of the just observable difference as that difference which will always be correctly appreciated but which cannot be lessened without forfeiting this distinction, we are simply making a special method of that instance of the method of right and wrong cases where (in a limited number of trials) the number of right answers equals the total number of answers. As already noticed, an error here and there is usually considered allowable; but if a few, why not more? Is there anything but the arbitrary preference of the experimenter upon which to base a decision? A more radical objection, however, remains to be noticed. Theoretically a difference of stimuli great enough to ensure a total avoidance of error must be infinite. With a certain ratio of stimuli one may as a matter of fact find no errors, but there is no (theoretical) assurance that if the experiments were sufficiently continued an error would not occur. This point is too obvious to need further illustration. We have seen that in its pure form the method of the just observable difference

simply measures the confidence, the disposition towards answering, and only with whatever accuracy the confidence can be experimentally proved to be a measure of the sensibility, has it value in determining the latter. In the form in which the method is generally used it is simply a loose and inaccurate application of the method of right and wrong cases. The method may be of service in determining roughly within what limits it is advisable to experiment, and has other obvious (practical) uses.

If the mischief to which this spurious method has given rise were confined to the charges already brought against it, the case, though serious enough, would not be as serious as it really is. In addition the method has given rise to radically wrong conceptions, chief amongst which is the conception of the threshold (*Unterschiedsschwelle*). This conception grew directly out of the method of the just observable difference; in fact this difference has by some been taken to be the differential threshold. What is more usually denoted by the threshold is the smallest difference that can be perceived. It is the threshold of consciousness. The moment we define this term accurately its unscientific character becomes apparent. The threshold is described as a point not exactly constant, but nearly so; above it all differences can be felt, below it all differences vanish into the unconscious. No matter whether little or much below this point, they are all utterly lost; it is idle to say, as Fechner at times does, that they differ in the amount of additional stimulation necessary to bring them up into consciousness, unless you mean that the series below the so-called threshold is an exact continuation of the series above it—and if you do mean this, then the threshold loses all its distin-

guishing peculiarities and ceases to exist. Either there is a threshold—be it a point or a more or less variable line—below which is homogeneous unconsciousness; or from the region in which the *sensed* difference has its maximum of clearness down to the point where it utterly vanishes because the difference between the stimuli vanishes, there is a continuous series of intermediate degrees of clearness, and there is no point on the curve with characteristics peculiar to itself, no threshold in any true sense.

But it will be well to postpone further consideration of this point until the nature of the method of right and wrong cases has been delineated, merely calling attention to the fact that, as Fechner and others admitted or even vaunted, the method of the just observable difference is the only one that is closely or at all connected with the threshold theory, and naturally the two may be discarded together. Sensation and stimulation each forms a continuum, and it leads to hopeless confusion to apply discrete conceptions to them.

CRITIQUE OF METHOD II.

The method of right and wrong cases is a device by which the sensibility can be determined while the judgment has but the simple problem of greater or less, of yes or no, to deal with. The factors of which this method makes use are the two stimuli of which the larger bears to the smaller a certain ratio to be known as $1 + x$; and the ratio of *wrong* answers to the total number of answers to be known as n . While the ratio $1 + x$ may have any value whatsoever, it actually does not differ much from unity, because only by the employment of such ratios can a number of right and wrong cases be readily collected. The all-important law,

justified by theory as well as by practice, announces that *ceteris paribus* as you diminish the difference between the two stimuli the number of wrong answers will increase, and as you increase the difference between the stimuli the number of wrong answers will diminish. In other words, as $1 + x$ increases n diminishes, and as $1 + x$ diminishes n increases. This law, that the number of wrong answers and the difference between the stimuli vary in opposite senses (the nature of this variation is not now in question), I regard as *a* if not *the* fundamental proposition of psycho-physics. (There are some conditions as to the nature of the experiments which must be complied with before the law will be found good; these will be assumed for the present and discussed later on.)

The all-important point is to decide how this "inverse law" is to be interpreted. In the first place, having found it true when the two stimuli differ by a quantity x , we expect to find (and will find) it true when they differ by $\frac{1}{2}x$. As, however, the judgments of the subject are under the influence of the many slight variations of condition that always influence psychological processes, it is possible that in a particular instance (especially when x or the total number of trials is small) the difference in the number of wrong answers will fail to appear. But if the number of experiments be increased, *and be increased the more the smaller the value of x* , this difference *will* appear. It must be remembered, however, that it may be impracticable to collect sufficient observations to bring out the more minute differences. But we have a right to infer that under proper conditions they would appear. If I find that as I successively experiment with stimuli that are related as 1 to $1 + x$ and then

with stimuli related as 1 to $1 + \frac{1}{2}x$, and then as 1 to $1 + \frac{1}{3}x$ and as 1 to $1 + \frac{1}{4}x$, the ratio of wrong answers, n , successively increases from one to the other, it certainly is in the highest degree improbable that the law does not hold with intermediate fractions of x . In other words, we infer that it can be expressed by a continuous curve, and that we must theoretically regard the probable ratio of wrong answers with two stimuli differing by the ratio x as smaller than the ratio of wrong answers with two stimuli differing by any fraction of the ratio x (1) in general, no matter what the value of x is, and (2) in particular, no matter how near zero that fraction of x is. Any one admitting these propositions (and it is not clear on what grounds they can be questioned) must logically endorse the psycho-physic reform which I am about to advocate; and I cannot but think that if psycho-physics had been built with a due consideration of these propositions, that science would have been a different and a sounder one. As a matter of fact all, even those least in sympathy with the point of view here taken, admit the law (1) in general when the value of x is confined within certain limits, and (2) in particular when the value of x does not too closely approach zero. Of course they have not stated their position in this obviously incorrect way; but if asked to state it they could not, as I understand it, state it in any other way. To show how totally without justification such a position is, it is sufficient to state that the choice of what values of x shall be admitted and what excluded, as well as of what limit is to be set to the lowest value of x , is and must be to a large extent an arbitrary one.

We can now return to the discussion of the theory of the threshold. What from the point of view of the

method of right and wrong cases does the current conception of a threshold demand? Nothing less than the position just now refuted. It has been proved that the ratio of wrong answers increases as the difference between the stimuli decreases; but the "threshold theory" claims that this law fails to hold after this difference has been diminished below a certain ratio. It actually says that you will oftener err in judging between weights of 30 and 32 ounces than in judging between weights of 30 and 34 ounces; oftener in judging between 30 and 31 ounces than between 30 and 32 ounces; but (supposing the so-called threshold to be $\frac{1}{30}$) that you will NOT err oftener in judging between 30 and 30.5 ounces, or between 30 and 30.1 ounces, than in judging between 30 and 31 ounces. Or, from a psychological point of view, it must propound the strange proposition that while under favorable conditions you will be enabled to appreciate the difference between 30 and 31 ounces, the conditions will never be sufficiently favorable to enable you to appreciate the difference between 30 and 30.5 ounces. Although the conception of the threshold is made highly improbable, and even irrational, by the consequences to which it inevitably leads, it is not necessary to be satisfied with a theoretical refutation. One can experiment with several differences all well below the limit assigned as the differential threshold, and actually find out whether or not the ratio of errors with the smaller of any two of these "sub-minimal" differences will be greater than that with the larger. Mr. C. S. Peirce and the writer undertook such a series of experiments (*v. Memoirs of the National Academy*, Vol. III), and found most conclusively that the law that the ratio of errors varies in an inverse sense with

the difference of the weights to be distinguished holds good as far as it is practicable to test it, and presumptively holds good throughout. It certainly holds good far below the limit assigned as the differential threshold for pressure, and if experiments could be sufficiently accumulated would be found good for still smaller differences. I do not forget that it would require an almost infinite series of experiments to make the most minute differences appear; just as at the other end of the scale one would "never" err in judging between 1 ounce and 12 ounces, and "never" in judging between 1 ounce and 13 ounces; yet it leads to more correct and practically useful conceptions to assert that an error is more probable in the former case than in the latter. This train of argument will be perfectly familiar to mathematically minded persons. (For a further discussion of the threshold *v.* Appendix A.)

We are now prepared to consider with more thoroughness the nature of the method of right and wrong cases. While, as the name of the method implies, the experiment is to be so arranged that the subject is to have the choice of two, and no more than two, answers, one of which shall be right and the other wrong, only a small number of experimenters have followed this rule. The violations of it have been of two kinds: first, in allowing the subject *three* answers (by using three kinds of pairs of stimuli, viz., having the first stimulus greater than the second; having it equal, or having it less); second, in allowing the subject the privilege of answering in one of these ways, or of saying that he is "doubtful." The objections to the first of these proceedings are evident and conclusive. As will be presently seen, one of the conditions on which the validity of the method of right and wrong cases depends is that the number of answers correct by the

action of chance shall be known ; and the possibility of three answers introduces an awkward and useless confusion into the calculation of these chances, and requires a much larger number of experiments to ensure equally reliable results. For in each case the answer given by the subject may be (1) correct, (2) singly wrong, and (3) doubly wrong ; *e. g.* if the first stimulus is really more intense than the second the subject may call it greater and his answer be correct, may call it equal and his answer be singly wrong, or may call it less and his answer be doubly wrong. The calculation of the chances of a correct, a singly wrong, and a doubly wrong answer by mere guesswork is certainly a very delicate one, especially when you consider that when the stimuli are really equal a doubly wrong answer is impossible and there are two ways of having a singly wrong answer. Again, with fine differences it will be difficult to keep the three kinds of sensations in mind, and slight lapses of the attention in such cases would favor the judgment that the stimuli are really equal. But objections to this proceeding could be indefinitely multiplied,¹ and one very curious one is given further on. Suffice it to say that it has no point whatever in its favor, and is really antagonistic to the spirit of the method upon which it foists itself. The method of right and wrong cases is a justifiable and a good one for measuring sensibility ; the method “ of right, wrong, or equal cases ” is certainly a different one and has no justification.

The objection to allowing doubtful answers is also apparent. If you do allow them, what are you to do with

¹A very serious objection is that it will be difficult, if not impossible, to take adequate account of the difference in sensibility for an increase and a decrease of sensation, which, it will be shown later, it is necessary to do.

them? Neglect them? Then when your x is very small you will have to throw out most of your experiments, and will in fact be recording only the best ones; while the very condition that makes the method of right and wrong cases a valid one is that *all* errors be recorded. Errors are due to slight lapses of the attention and all the other fluctuations to which the judgment is subject; to allow doubtful answers is to rule out all cases in which the judgment is in a somewhat worse than its average condition, and thus to vitiate the real average. It would be quite as justifiable (in fact, if you do the one you ought to do the other) to rule out all cases in which the subject feels *unusually certain* of the correctness of his answer. But may we not, as Fechner¹ and many others did, count half of the doubtful answers right and half wrong? Certainly not. (1) Because all judgments must be recorded as given; (2) because that would give a fictitious appearance of having made more observations than you really have, and when x is small (and the number of doubtful answers large) would seriously influence the meaning of the result; (3) because, while it is true that the chance of any answer in general being correct is one half, it is not at all likely that the chance of this particular kind of an answer being correct is as much as one half. Other more practical objections to the process are that it encourages fatigue and diminishes the regularity and simplicity of the judging process. In fact the only point in favor of this proceeding is that this doubtfulness is a real and valuable symptom; but a truer and

¹ Fechner used this method at first, but gave it up later. It is to be noted that if you allow "equal" answers you *certainly and especially* cannot count the doubtful answers half wrong and half right, because in that case the chance of a right or wrong answer is not one half.

better mode of taking this into account is described under the term "confidence" in Appendix B.

We are now able to define the method of right and wrong cases, and having done so, may pass on to its theoretical justification. The method may be formulated thus: Having chosen two stimuli of which the one bears to the other the ratio¹ $(1 + x)$, apply them to the sensitive surface of the subject, requiring him to decide in each instance whether the first stimulus is greater or less than the second (one of these answers being right and the other wrong, so that by mere guesswork he will answer correctly in one half the cases), and record the ratio (n) of wrong answers to the total number of answers.

As already stated, the errors may be regarded as due to lapses of the attention, slight fatigues, and all the other numerous psychological fluctuations that go to make us now better and now worse judging agencies than our average selves. These influences may be said to have the effect of loading the smaller stimulus (or lightening the larger) so as to come up to (and, strictly speaking, just overtop) the greater stimulus. As long as $(1 + x)$ remains constant, the "amount of work" which these accidental fluctuations must perform in thus loading the smaller stimulus sufficiently to cause an error remains constant, and is really x ; hence, as $1 + x$ diminishes this work diminishes, and as $(1 + x)$ increases it increases. These fluctuations, it is understood, are of such a nature as to be as frequently and to the same degree in favor of our judging powers as antagonistic to them; and the probability of their accumulating sufficiently in one direction to cause an

¹ $(1 + x)$ and not x is chosen to show that the ratio of the larger to the smaller stimulus is designated; this choice is made on grounds of convenience only.

error when the first stimulus bears to the second the ratio $(1 + x)$ is less than the probability of their doing so when the ratio of the two stimuli is less than $(1 + x)$, and is greater than when that ratio is greater than $(1 + x)$ —*i. e.*, with a smaller $(1 + x)$ errors are more probable and hence will occur more frequently than with a larger $(1 + x)$. And the law that regulates the probabilities of the deviations by various degrees from the average (*i. e.*, the frequency of error with various stimulation ratios $(1 + x)$) is the law expressed by the “probability curve,” which pictures the effect of a very large (strictly infinite) number of small causes no one of which has of itself any decided influence. And here we have touched bottom. This law forms the basis of the method of right and wrong cases, and enables us to predict what ratio of errors will occur with any value of $(1 + x)$ when we have experimentally determined in a given case this ratio, n , for a given value of $(1 + x)$.¹ The formula for doing this, together with illustrations of its application, is given in Appendix C.

As the object of any psycho-physic method is to measure the sensibility, it remains to show how this is to be done, and thus to supply the want which the just observable difference was intended to meet: namely, to afford a ready method of comparing the variation of sensibility in different individuals, at different times, with different modes of judging, in

¹ The paper by Mr. C. S. Peirce and the writer, above referred to, illustrates the close correspondence between theory and practice in this respect. The only difficulty in showing this arises from the fact that the probable error is constantly decreasing (due to practice and so on), and one must therefore divide the results into groups within which the probable error is presumptively tolerably constant. This consideration must also be taken into account in calculating an n for a certain $1 + x$, on the basis of several n 's obtained with several $(1 + x)$'s.

different senses, and so on. The frequency of error with each value of the ratio $(1+x)$ is expressed by a continuous curve when $(1+x)$ changes gradually; there is no characteristic point on the curve evidently appropriate for the standard of sensibility, hence the choice of such a point must be made on grounds of convenience and simplicity. The standard of sensibility that I now propose is that ratio of the two stimuli (or rather that ratio less one) with which one half of the answers being correct by chance, one half of the remaining one half of the answers will also be correct—*i. e.*, when *one* error occurs in every *four* answers. The reason of this choice is that this ratio measures the probable error, or that error which is as likely to be exceeded as to be fallen short of. This will be fully explained in considering the method of the average error, and it will there be shown that this standard of sensibility forms the easiest possible transition between the method of right and wrong cases and that of the average error; which, I take it, is an essential requisite of a standard of sensibility. Of course it is not necessary that this value be experimentally found; it is to be calculated by the formula given in Appendix C, from any ascertained ratio of error with any ratio of stimuli. If several such data are at one's disposal it is to be calculated from each, and the mean drawn or treated in what is recognized as the fairest manner.

Having thus obtained a standard of sensibility, it only remains to illustrate its application and to mention some practical conditions which this method makes advisable. I will call that ratio of excitation with which errors occur once in four times the "*standard ratio*." If, for example, I find as the

result of 1000 experiments with two weights 200 and 210 grams, that 250 of the answers are wrong (or calculate from an equivalent set of experiments that at this ratio of stimuli that proportion of answers would be wrong), then the sensibility of the pressure sense in this case is $\frac{1}{20}$. (The stimuli being 200 and 210, $1+x = \frac{210}{200} = 1 + \frac{1}{20} \therefore x = \frac{1}{20}$). If in a following series of experiments I find 250 mistakes when the stimulation values are only 200 and 208, then the sensibility has improved from $\frac{1}{20}$ to $\frac{1}{25}$; and thus the sensibility is said to be twice as fine (not when half the ratio of errors are made, but) when the ratio of stimuli necessary to produce the same ratio of errors is halved. If in a given case I find that A makes fewer errors than B, I have only to calculate the standard ratio for each in order to quantitatively ascertain the ratio of their sensibilities, which are inversely proportional to their standard ratios. And finally, I can compare different senses on the generally admitted supposition that their sensibilities are to be measured by the *ratio* of the stimuli (apart from their absolute value) leading to equal ratios of error. This comparison would lose much of its significance and validity in case Weber's law does not hold; for then no such method of comparing entirely different senses would exist, inasmuch as the absolute value of the stimulus would then be important and there is no connection between an ounce and an inch. If Weber's law is true within limits, the comparison holds only within those limits.

If the plan of experimentation thus far sketched were followed there would be little room for serious error, and the experiments of various observers would be generally comparable. I will, however, add some suggestions and precautions, all of which have proved themselves highly advisable, if not essential.

(a). It should be stated what knowledge the subject has of the conditions and purposes of the experiment, and the subject should know all the conditions except such as will lead to the use of indications towards forming a judgment other than those furnished by the sensation itself. If he is in doubt as to the several changes that can possibly occur he will infer them for himself, and will yield to that uncontrollable psychological guessing of what is coming. This mischievous tendency plays havoc with the expectation and throws the attention off the track. When the confidence is low the tendency to prefer one kind of answer is apt to occur, and would be avoided if the subject knew that this tendency had no basis in fact. It is especially necessary for the subject to know that in each case the one stimulus is greater or less than the other (and never equal to it), and that either is as liable to come first as last.

(b). The greater sensibility for an increase than for a decrease of sensation must be taken into account. This simply means that as a matter of fact one is more apt to perceive a change from 20 ounces to 21 ounces than a change from 21 ounces to 20 ounces, and therefore the two experiments should not be placed on a par. This caution is quite usually observed ; but what seems to me the easiest method of avoiding the difficulty is used, as far as I know, only in the experiments on "Small Differences of Sensation" above referred to. It consists in having one of each kind of change in each experiment. For example, in the above case the order of the weights would be (1) 20, 21, 20, or (2) 21, 20, 21, the subject being required to decide whether the middle stimulus was greater or less than the first and third stimulus. The mean of two sets in one

of which only "increases" and in the other only "decreases" are used is very good, but multiplies the number of experiments without in general yielding any compensating advantages. Cases may arise, however, in which the first mentioned process is inapplicable; but these are rare, and in general the "double process" is advisable. It gives two chances of judging and makes the conditions highly favorable. It should be distinctly stated which method is used. In either method one half the answers will be right by the action of chance.

(c). It is highly advisable to have as many of one kind of change as of the other, *i. e.*, as many "decreases" followed by "increases" as *vice versa*, or, if the other method is used, as many "decreases" as "increases." This is generally conceded; and the only point worthy of mention is that one can avoid the subject's taking any unfair advantage of this fact by having a large number of experiments in one set. If that is impracticable, divide a large set composed of an equal number of each kind of change by a *chance* arrangement into smaller groups. In general let a *chance arrangement* (die throwing, etc.) decide the order of the several kinds of changes.

(d). A precaution that I have found of great value is that the moment at which the change is to occur shall be under the control of the subject, and not, as is usual, at the command of the experimenter. In this way the subject knows exactly *when* to expect the sensation, and he can ask for it at the moment when he is best prepared to receive it. Any slight non-distracting movement can be agreed upon as the signal to mean "change." This is a greater advantage than would at first sight appear.

(e). It is hardly worth while adding that the method should be the same throughout, that the conditions be kept as equable as possible, that the effect of practice be noted and of fatigue avoided, and so on and so on.

It will be worth while illustrating by a single example, to what kind of work the employment of a wrong method leads. The author in question is experimenting with the pressure sense.¹ With a constant initial weight of 10 grams he successively increases the differential weight from .1 gram up to the point where, in a set of 16 trials, no wrong answers occur. He thus uses the so-called method of the just observable difference in its worst form, and arbitrarily fixes the point of no error at no error in 16 answers. Not satisfied with this, he in some cases allows an error or two and still calls it the just observable difference. Again, he has three pairs of changes; beginning with 10 grams he either (1) increases, (2) repeats it (equal), or (3) decreases it. He thus makes it impossible to know the number of answers correct by chance. The answer when the change was an increase or decrease may be correct or doubly wrong (*conträr*); when the weights are equal it may be correct or wrong (*falsch*). What happened when the subject said the weights were equal but they were not so is not recorded. Finally, the subject could always say, if he chose, that he was undecided (*unbestimmt*). The object is now to increase the differential weight until only "correct" answers remain, or nearly so. With differences of .1, .3 and .5 gram the *whole table* conveys the important information that the subject never felt like answering at all. With .4 gram he begins to answer,

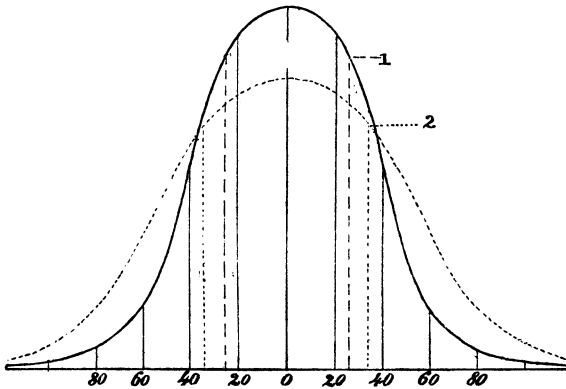
¹ *Experimentelle Prüfung der zu Drucksinn-Messung angewandten Methoden*, etc. Von Dr. Bastelberger. (Eine von der Universität Strassburg gekrönte Preisschrift.) Stuttgart, 1879, pp. 70.

frequently makes mistakes, and at last when the 10 grams are increased by 4.5 grams, or the ratio is nearly 2 : 3, he commits no errors in a set of 16 trials. In this process there were 192 experiments, in *only* 68 of which did the subject give a definite answer. Up to the very last "doubtful" answers occurred. Results like this are utterly unreliable. They show nothing, and are not comparable with the results of others. What has been done is to show that in a particular set of 16 experiments the subject was disposed to answer each time and made no errors, and that with small differences the confidence is low. It is certainly not to be supposed that if the experiments were repeated one would get the same number of doubtful, of right, of singly wrong, and of doubly wrong answers, with the same ratio of stimuli ; and if they differ, we may not be able to tell whether the change is an improvement or a deterioration. Again, the results say that when I put 10 grams on the finger and then take it off and replace it I will not know that the same weight has been replaced, if this occurs in a series when the greatest change to be expected is an addition or diminution of .1 gram, but that I will recognize this *very same* act when it occurs in a series in which changes of .4 gram are made. This difference can only be due to contrast with the preceding and the expected sensations, and is therefore an effect which interferes with an accurate determination of sensibility. And all this is simply the effect of allowing "equal" answers. In short, a correct method would yield results which included these and said much more in half the time.

CRITIQUE OF METHOD III.

We come at last to the most natural and in some respects the most important method, that of the average error. This, like the former, is founded upon the probable error ; in fact, the x of the standard ratio is the probable error. To show what elements are involved in this method and how they are dealt with, let us take a particular instance. Let the problem be to draw lines equal in length to a given line ; in so doing the average result will be (1) to draw a line really equal to the original line, *i. e.* the several exaggerations and under-estimations will balance one another ; or (2) to draw a line slightly longer, or (3) one slightly shorter, than the original line. In either case let the point marked zero represent the average result of the reproductions ; the points 20, 40, 60, etc. to either side denoting reproductions differing from the average by 20, 40, 60, etc. per cent of the average respectively. Now in any extended series of such experiments the number of reproductions of nearly the average length will be larger than the number of reproductions differing more from the average. And the law which this grouping about an average follows is that expressed by the probability curve. In other words, that curve pictures the frequency of each degree of fluctuation to which the judgment is exposed, the abscissae measuring the extent of the error in each direction, and the ordinates the ratio of errors of each degree of error. The *average error* of these adjustments is obtained by dividing the sum of the deviations from the average (without regard to sign) by the total number of adjustments ; the *probable error* is obtained by simple mathematical processes (explained in Appendix D, *q. v.*), and measures the

limits within which any observation is as likely to fall as it is to fall beyond them. It is that point on the curve the ordinate from which divides each half of the curve into two equal areas, and is thus represented by the dotted ordinates.



This probable error is the gauge of variation in sensibility from day to day, in different individuals, and so on. The point which it marks is chosen mainly for convenience and the simplicity of the formula to which it leads.

Let me illustrate how the probable error measures the sensibility. If to-day I am in a better judging condition than I was yesterday, my probable error will be less ; this means that I will be more consistent, be less subject to large disturbing variations, and react more nearly in the same way on each occasion. This is what is indicated by the probable error, and (as we saw in the method of right and wrong cases) is what we do and ought to mean by being able to judge better. Again, if A is a better observer than B, the complete significance of this fact is expressed by saying that his

probable error is less than B's ; this is fully recognized in astronomical and other exact observations. It is not an ultra-refinement, but is at once the simplest, most accurate and adequate mode of expressing those differences with which psycho-physics deals. Finally, to say that one sense is finer than another is to say that its probable error is less. For example, the sense of vision is finer than the pressure sense ; this means that if I repeatedly select from a large number of slightly different weights one that shall equal a given weight, the point in each half of the curve which has an equal number of errors to either side of it will be farther off in the pressure curve than in the curve resulting from matching two lines by the sense of vision, as shown in the figure. Moreover, the probable error furnishes a quantitative estimate of sensibility. If A has twice the sensibility of B this means that his probable error will be one half that of B. If the effect of practice is to increase my sensibility by one half its first amount, my probable error will decrease by one third of its amount, ($\frac{1}{1.5} = \frac{2}{3}$; $1 - \frac{2}{3} = \frac{1}{3}$).

Let us return for a moment to the three possible results of the method of the average error (or, as we have just seen reasons for terming it, the probable error) as above given. If the average result of all the adjustments equals (or nearly equals) the real intensity of the first stimulus, it shows that the causes leading to error in one direction are equal in efficiency to those causing errors in an opposite direction. If, however, the average result of all the adjustments shows a constant deviation (either greater or smaller) from the original stimulus, then there is a constant and a variable error, which two are totally different and independent things. The constant error must be ascribed

to some peculiarities of our organism and so on, and has no value whatever in measuring the sensibility ; this, as before (and always), is measured by the probable error of the deviations of the several adjustments from their mean. The constant error measures something very important and forms a special object of research. But, it will be asked, how will this constant error appear in the method of right and wrong cases ? As regards the ratio of error it will not appear at all. This constant error would appear in the curve as a shifting of the central axis to one side. This does not affect the probable error, which alone decides the ratio of error in the method of right and wrong cases. Why it does not thus appear may be seen from the following considerations. The constant error makes the probability of a certain deviation—inasmuch as that deviation is made larger by the existence of the constant error—*less* than if no such error existed; but this is exactly counterbalanced in those equally frequent cases in which the constant error aids to the same extent in *lessening* the degree of a deviation. For example, if the effect of the constant error is to lead me to regard a line $\frac{5}{4}$ of the first line in length as its equal, then in the method of right and wrong cases this means that the probability of my making an error of any degree is made less because I must now make an error $\frac{5}{4}$ of its size ; this is when an increase is taken to be a decrease. But when I mistake a decrease for an increase, the additional $\frac{1}{4}$ to the length of the line by that much *decreases* the size and increases the probability of such an error. But if the effect of the constant error is such as to always make the altered (not the initial) stimulus seem larger, then the constant error will appear in the fact that more errors in taking a decrease to be an

increase than *vice versa* will occur ; and in fact we can quantitatively determine the constant error by taking half the difference between the probable error of all the judgments of one kind and that of the other kind of judgments—a proceeding which I do not remember to have seen in practice.

There are no special precautions necessary in carrying out this method. It is natural and easy, but not practically applicable to all senses. One must take care that the subject really has a free choice of all such reproductions as he is at all likely to choose.

CONCLUSION.

It will have been noticed that this critique has dealt solely with the theoretical and practical justifications of the three usually recognized psycho-physic methods. It has avoided any reference to the psycho-physic law in Fechner's sense, and only in a few places has it been led to consider Weber's law. Weber's law is either (1) true throughout the psychic scale, or (2) it is not true at all, or (3) it is true within limits. In the second case, as has been noticed, we lose a valuable method of comparing the accuracy of different senses unless a law similar to that formulated by Weber can be proved to hold. In the third case we must limit our comparison of different senses to those absolute stimuli which show the greatest tendency to be in accord with Weber's law. The question of a practical correction for the lower and upper end of the sensitive scale in each sense is a separate one, and cannot be considered here. My object now is to point out that a main function of Weber's (or any similar) law is to supply a method of comparing the sensibility of different senses, and the

function of the two legitimate psycho-physic methods is to furnish standards of sensibility in the several senses. The results obtained by either of these methods can be expressed in terms of the other. The difference between them is in the psychological processes of which they make use ; and it is possible that this difference is so great as to some slight extent to vitiate the mathematical relations that have been deduced for transition from one to the other. This can only be decided by actual experiment ; and such experiments, if sufficiently numerous and carefully conducted, would form a valuable contribution to the subject. If the result were to show an agreement between theory and practice (as I believe it would), it would give an especial significance to the definition of man as a rational animal.

Finally, a word as to Fechner's law, which reads that the sensation is proportional to the logarithm of the excitation. That law in one sense, I believe, can be deduced from Weber's experiments only by the use of a series of assumptions, hardly one of which is even probably justifiable. Fechner has confused "the sensation of being different" with "the difference of sensation," and his law seems to me, in the sense in which it is often, if not usually, stated, to be without truth or meaning. But I reserve all criticism of this as well as of other fundamental propositions in the logic of psycho-physics for another occasion, and will conclude this paper with a summary of the main points which have been advanced therein.

(1). The method known as the method of the just observable difference is either not at all suitable for an exact measurement of sensibility, or it is but a loose application of the method of right and wrong

cases. It should therefore be omitted from the psychophysic methods, where it has introduced much confusion and many misconceptions.

(2). The threshold is such a misconception, arising from a discrete mode of regarding continuous quantity ; and is as valueless as a standard of sensibility as it is unjustifiable theoretically. The variations of the probable error form a continuous curve, while the threshold theory requires a more or less sudden change in the direction of this curve.

(3). The method of right and wrong cases is justifiable when used with certain precautions ; in particular, when but two answers are possible and but two kinds of excitation are used ; when the subject is required to record a definite answer each time ; when the number of answers correct by chance is known (and equals one half). Other advisable rules are given in the text.

(4). The justification of this method lies in the fact that the causes of error follow the probability curve ; and thus a means is furnished of calculating either the ratio of errors at any given ratio of stimuli, or the ratio of stimuli at any given ratio of error, when the ratio of errors at any one ratio of stimuli is known.

(5). The standard ratio by which sensibility is to be measured is that ratio of stimuli at which one error occurs in every four answers.

(6). The method of the average error (better, of the probable error) depends directly on the ascertaining of the probable error ; and the probable error itself measures the sensibility. The x of the standard ratio in the method of right and wrong cases is the probable error, and this fact yields a ready method of comparing the results of the two methods.

(7). The function and value of Weber's law depends

on its furnishing (it may be within limits) a means of comparing the sensibility of different incommensurate senses. It can be formulated in terms of the method of right and wrong cases, as saying that the standard ratio is independent of the absolute value of the stimuli but depends solely on their ratio $(1 + x)$; and in terms of the method of the average error, as saying that the probable error will be uninfluenced by a change in the absolute size of the stimulus according to which the adjustments are to be made.

APPENDIX A.

The Practical Threshold.

While I maintain that the theoretical refutation of the threshold theory and the establishing of the point of view of the probable error carries with it the assurance that no practical difficulty to which they may give rise will be more than an apparent one, yet it may be worth while showing how such objections are to be met. The favorable evidence which the assumption of a threshold derives from ordinary experience can be illustrated thus: We do not see the stars at day, yet they are there. This can only be because the lustre added by their brightness to the enormous sunlight already existing is too insignificant ever to appear visible to our eyes; it is lost below our differential threshold. In so extreme an instance the difference between the current view of the threshold and the one here advocated becomes theoretical only; but that does not lessen its importance. Consider the facts more closely; at day the star is invisible, at night it is visible. Hence, the argument reads, there must be a point where the visible passes into the invisible at dusk and comes back into the visible again

at dawn. The question is, what is the correct mode of describing this process. The current method is this : the ratio of the brilliancy of the star to the already existing light is constantly increasing, and when this ratio has increased beyond a certain amount (the differential threshold for vision) the star becomes visible. My explanation would be this. I would first call attention to the fact that the star would be invisible to some persons when it is visible to others, would under parallel conditions be invisible to me one day at a given time and visible the next day, in order to show that the term threshold is intended to refer to an average threshold. I would then ask whether you will always be able to see the star a minute time after the ratio of its brilliancy to that of the sun has increased above the ratio referred to. If you answer "yes" you define your threshold to mean that ratio of the brilliancy of the star to the sun at which all your answers will be correct. Here you either (1) tacitly assume that not many observations are to be taken, or that (2) no matter how many observations were made no mistake would ever occur. If you mean the former you admit that if the observations went on errors might occur ; but the causes which led to these errors have not totally vanished, but have only gradually decreased without any sudden break in the process—*i. e.* without any threshold. If you mean the latter you are claiming a very improbable proposition ; for the causes leading to error still exist, and though very minute, and errors rare, still they are never impossible. *Practically* they will be impossible after a certain more or less definite point ; but this simply means that it would be impracticable to collect sufficient observations to ensure the occurrence of an error. One can agree to mean by a *practical* threshold that ratio

of excitations at which no more than one in a hundred or one in two hundred answers will occur—*i. e.* one can agree to neglect all causes of error not sufficient to produce at least one error in one or two hundred trials; and can use this as a standard ratio, to be calculated as the other standard ratio. But reasons have been given for preferring the standard ratio first proposed. If, however, a practical threshold be desired, it can be agreed upon, but it will not be a real threshold in any true sense. The star would be far below such a practical threshold.

I am indebted to Dr. Fabian Franklin, of Johns Hopkins University, for pointing out that there is a form of the threshold theory consistent with the mathematical basis here advocated. It is this: we can imagine a ratio of stimuli differing very slightly from unity which a judgment less subject to fluctuations than ours would (owing perhaps to some peculiarities of its organism) more often disregard. And the more perfectly free from fluctuations such a judgment is, the more automatic the process of judging, (not the more often will this small difference be perceived but) the more often will it *just fail* to be perceived. We are dealing not with more and more observations but with a better and better judgment. And as this judgment approaches perfection we can imagine it perfectly perceiving certain differences and perfectly failing to perceive all differences below a certain fixed difference, which would thus be the threshold.

In reply to this I have only to state that (1) from the experience that we have we can assert that such a state of things is extremely improbable, and (2) that if it were true it would necessitate the same psycho-physic methods which are here considered valid, and that in

brief the practical outcome of it would be quite the same as those that arise from the theory here advocated. It might be worth while devising experiments to test the possibility of this supposition. It is to be noted that this form of the threshold theory is as antagonistic to the old threshold theory as the one advocated in this paper. Such a threshold must be very much more minute than any value assigned as the differential threshold in the old sense.

APPENDIX B.

The Method of Gradual Increment.—The Confidence.

In discussing the method of the just observable difference it was implied that though the usual method of that name was not valid, there was a genuine form of the method. In its true form it has recently been applied to the study of the pressure sense.¹ It consists in allowing the initial weight to change *gradually*, and to find *when* the subject detects the direction of the change—whether an increase or a decrease of pressure.

A study of the nature of this proceeding sheds much light on the operations involved in the process of judging. The sensation gradually changes, and the question is how soon is this change detected? In the first place it is to be noted that there are two variables, the rate of change and the amount of change. For the sake of simplicity suppose the rate of change constant. By how much must my sensation change before I am willing to decide in what direction it has changed? My point is that this is to a large extent an individual matter. It means what is the smallest amount of confidence upon which I will risk a judgment. If I wait until I feel perfectly certain about it

¹ See the article by Hall and Motora in No. 1 of this Journal.

my "just observable difference" will be large; if I judge as soon as I have a minimal amount of confidence I will have a small "just observable difference," but will doubtless make many mistakes. This feeling of confidence is what the "just observable difference" method takes into account. And we would expect that the ratio of errors in the method of right and wrong cases varies in an opposite direction not only with the difference judged, but also with the confidence in the correctness of that judgment. When the difference of the stimuli is constant the number of errors in the various sets will vary "inversely" as the confidence; and hence this subjective feeling may be utilized for recording the differences between individuals and between different series of judgments of the same individual. The feeling of confidence will itself be liable to variations, but every one will doubtless have a tolerably constant "index of confidence."

We see thus that in its true form the method of the just observable difference measures the disposition to answer, and this in turn is determined by the subjective feeling of confidence. The method is calculated to shed much light on the subjective states that accompany the act of judging, but though valuable in other directions, is not suited for measuring sensibility. It is also to be noted that as a considerable variation in the confidence from time to time, or even between different individuals, is not to be expected, that feeling of confidence which prompts one to answer may be considered sufficiently constant to enable one to base a rough measurement of the sensibility upon it and not upon the correctness of the answer. And in this way the just observable difference as ordinarily tested may be useful in hurriedly

testing sensibility, in pathological cases and elsewhere. What I have said is not opposed to such a use of it.

To record the confidence is a difficult and must to a large extent be an arbitrary matter. In the experiments on "Small Differences of Sensation" we used the following plan: 0 denoted the absence of any preference for one answer over its opposite, 3 denoted as strong a confidence as one would have in ordinary sensations, and between the two 1 and 2 naturally found their places. From records made on this plan Mr. Peirce deduces the formula $m = c \log \frac{p}{1-p}$, where m denotes the degree of confidence, p the probability of the answer being right, and c a constant which may be called the index of confidence. This formula closely approximates the results actually obtained. It appears, too, that with increased practice the index of confidence rises.

It was above deduced that the confidence must vary in a direct sense with the ratio of the stimuli, and in an opposite sense with the ratio of errors. This is very clearly shown in our experiments. Mr. Peirce's average confidence was .67 when the two stimuli were 1000 and 1060 grams; was .28 when they were 1000 and 1030 grams; and was .15 when they were 1000 and 1015 grams. Similar numbers for myself are .90, .51 and .30; and when the ratio of the stimuli was further diminished my confidence was still further reduced. With the stimuli 1000 and 1005 grams it was practically zero. In a paper published by the writer in *Mind*, No. 44, similar results are shown.

Again, the confidence varies in an inverse sense

with the number of errors. In Mr. Peirce's case only 3 per cent of all answers given with a confidence 3 were wrong; 10 per cent of those with a confidence 2; 18 per cent of those with a confidence 1; and 38 per cent of those with a confidence 0. Similar numbers for myself are 3, 6, 16 and 30 per cent. In spite of the obvious arbitrariness and inadequacy of this method it has proved itself surprisingly useful; it ought, however, to be improved in future work.

APPENDIX C.

Formulae for the Method of Right and Wrong Cases.

For these formulae as well as for important suggestions in several parts of this paper I am indebted to Dr. Fabian Franklin.

I. Rule for calculating the ratio of the two stimuli at which one fourth of the answers will be wrong when the ratio of wrong answers at any one ratio of stimuli is given.

Let $(1+x)$ be the given ratio of stimuli; let $(1+p)$ be the ratio at which one in four of the answers will be wrong; and let n be the ratio of errors with the ratio of stimuli $(1+x)$. The formula is

$$\log(1+p) = \frac{.477 \log(1+x)}{\sigma^{-1}(1-2n)},$$

in which $\sigma^{-1}(1-2n)$ means the t in a table of σt corresponding to $\sigma t = 1-2n$. Such a table is here appended and is taken from the article on Probability in the Encyclopedia Britannica, 9th edition. (It is to be noted that as the logarithms appear finally as a ratio they may be taken in any system of logarithms.)

Example 1.—In distinguishing between what weight and 100 ounces would A answer wrongly once in four

times, if he makes 15 errors in 100 answers when distinguishing between 100 and 105 ounces.

$$n = .15; 1 - 2n = .7; \theta^{-1}(1 - 2n) = \theta^{-1}(.7) = .7345; \\ 1 + x = 1.05; \log(1.05) = .0212.$$

$$\log(1 + p) = \frac{.477 \log(1 + x)}{\theta^{-1}(1 - 2n)} = \frac{.477(.0212)}{.7345} = .01377;$$

$1 + p = 1.032$. *Answer:* Between 100 and 103.2 ounces.

Example 2.—If in distinguishing between the brightness of two screens, the illumination of one of which is brighter by $\frac{1}{50}$ than the illumination of the other, B errs on the average 19 times in a set of 50 observations; what ratio of brightness must the second screen bear to the first for B to make only 12.5 wrong answers on the average, in a set of 50 observations?

$$n = .38; 1 - 2n = .24; \theta^{-1}(1 - 2n) = \theta^{-1}(.24) = .2164; \\ 1 + x = 1.02; \log(1.02) = .0086.$$

$$\log(1 + p) = \frac{.477 \log(1 + x)}{\theta^{-1}(1 - 2n)} = \frac{.477(.0086)}{.2164} = .01896.$$

$\therefore 1 + p = 1.0445$. *Answer:* The ratio $\frac{209}{200}$ (nearly).

II. One can find the ratio of stimuli at which *any ratio of errors* will occur when the ratio error with a given ratio of stimuli is known, by the following formula:

With the formula $\log(1 + p) = \frac{.477 \log(1 + x)}{\theta^{-1}(1 - 2n)}$ find

the value of p ; then with this value of p and the designated new value of n , find the value of $(1 + x)$ by the same formula transposed, viz.

$$\log(1 + x) = \frac{\log(1 + p) \theta^{-1}(1 - 2n)}{.477}.$$

Example 3.—In Example 1, with what ratio of stimuli

will A make only 7.5 errors in the average set of 100 observations ?

We had $p = .032$ or $\log(1+p) = .01377$, and $\theta^{-1}(1-2n)$ will now be equal to $\theta^{-1}(.85) = 1.02$.

$$\log(1+x) = \frac{(.01377)(1.02)}{.477} = .0295, \text{ and } (1+x) = 1.07.$$

Answer : 1.07.

III. One can also find the ratio of error at *any* ratio of stimuli when the ratio of error with one ratio of stimuli is given, by the following formula. Find $\log(1+p)$ as before. Then find n in the following formula where $(1+x)$ represents the new ratio of stimuli,

$$n = \frac{1 - \theta \left\{ \frac{.477 \log(1+x)}{\log(1+p)} \right\}}{2}$$

Example 4.—Find the ratio of wrong answers in Example 1 when the ratio of stimuli is 1.1.

We have $\log(1+p) = .01377$; $\log(1+x) = \log(1.1) = .0414$. Hence

$$\begin{aligned} n &= \frac{1 - \theta \left\{ \frac{.477 \log(1+x)}{\log(1+p)} \right\}}{2} = \frac{1 - \theta \left\{ \frac{(.477)(.0414)}{.01377} \right\}}{2} \\ &= \frac{1 - \theta(1.43)}{2} = \frac{1 - .956736}{2} = .021632. \end{aligned}$$

Answer : 2.16 errors in 100 answers.

IV. I will also show how a practical threshold can be obtained if desired.

Example 5.—Taking the practical threshold at one error in 100 answers and the probable error within its extreme limits in the case of the writer in the experiments on pressure above referred to, viz. .05 and .016, at what ratio will this threshold occur ?

$$\log (1+x)=\frac{\log (1+p) \theta^{-1}(1-2 n)}{.477},$$

$$\log (1.05)=.0212 ; \log (1.016)=.00689 ;$$

$$\theta^{-1}(1-2 n)=\theta^{-1}(.98)=1.649,$$

$$\log (1+x)=\frac{(.0212)(1.649)}{.477}=.07333 ; \therefore (1+x)=1.184,$$

$$\log (1+x)=\frac{(.00689)(1.649)}{.477}=.02383 ; \therefore (1+x)=1.056.$$

Answer : 100 ounces and 105.6 ounces in the first case ; 100 and 118.4 ounces in the second case.

Table of θt from $t=0$, to $t=3.0$.

t	θt	t	θt	t	θt	t	θt
0.00	0.00000	.2	.22270	1.3	.93401	2.4	.99931
.01	.01128	.3	.32863	1.4	.95229	2.5	.99959
.02	.02256	.4	.42839	1.5	.96611	2.6	.99976
.03	.03384	.5	.52050	1.6	.97625	2.7	.99986
.04	.04511	.6	.60386	1.7	.98379	2.8	.99992
.05	.05637	.7	.67780	1.8	.98909	2.9	.99996
.06	.06762	.8	.74210	1.9	.99279	3.0	.99998
.07	.07886	.9	.79691	2.0	.99532	∞	1.00000
.08	.09008	1.0	.84270	2.1	.99702		
.09	.10128	1.1	.88020	2.2	.99814		
.1	.11246	1.2	.91031	2.3	.99886		

Note.—Intermediate values in this table are derived by interpolation in the ordinary way.

APPENDIX D.

Rules for Computing the Probable Error.

These rules I take from Jevons, Principles of Science, p. 387.

1. "Draw the mean of all the observed results.
2. Find the excess or defect, that is, the error in each result from the mean.
3. Square each of these reputed errors.
4. Add together all these squares of the errors, which are of course all positive.

5. Divide by one less than the number of observations. This gives the *square of the mean error*.

6. Take the square root of the last result; it is the *mean error of a single observation*.

7. Divide now by the square root of the number of observations, and we get the *mean error of the mean result*.

8. Lastly, multiply by the natural constant 0.6745 (or approximately by 0.674, or even by $\frac{2}{3}$), and we arrive at the *probable error of the mean result*."

For illustrations of this process and methods for shortening the work see Jevons and works on "Probabilities" there referred to.

It is generally advisable to divide up the observations, and find the probable error of each group and then draw a mean. It is also sometimes desirable to be able to test how closely the number of errors of each degree of deviation from the mean follows the number assigned by the probability curve. Mr. Francis Galton gives an admirable account of this in an appendix to his "Hereditary Genius," to which the reader is referred.